Introducing scientometrics in the CORE Repositories Dashboard: a proposal

Petr Knoth & Nancy Pontika

CORE

The Open University

@oacore

What is CORE

What is CORE

Facts

> 192 API users

Facts

> 125 Repositories

Dashboard users

Facts

> 1034 Repositories

Facts

> 8,900 Journals

Facts

> 53 Languages

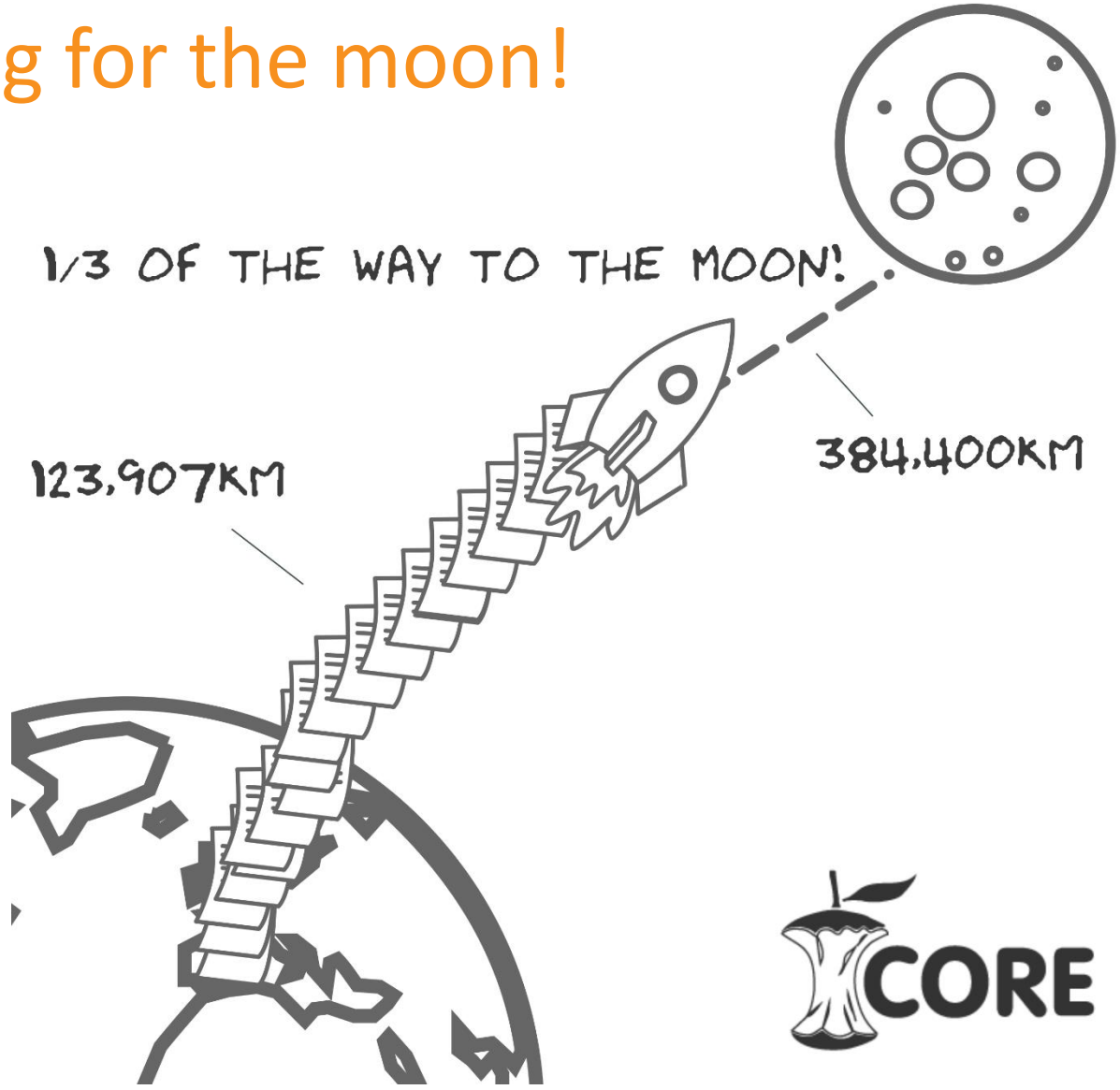
Facts

> 36,207,179 Metadata

Facts

> 3,800,995 Full-text

Aiming for the moon!



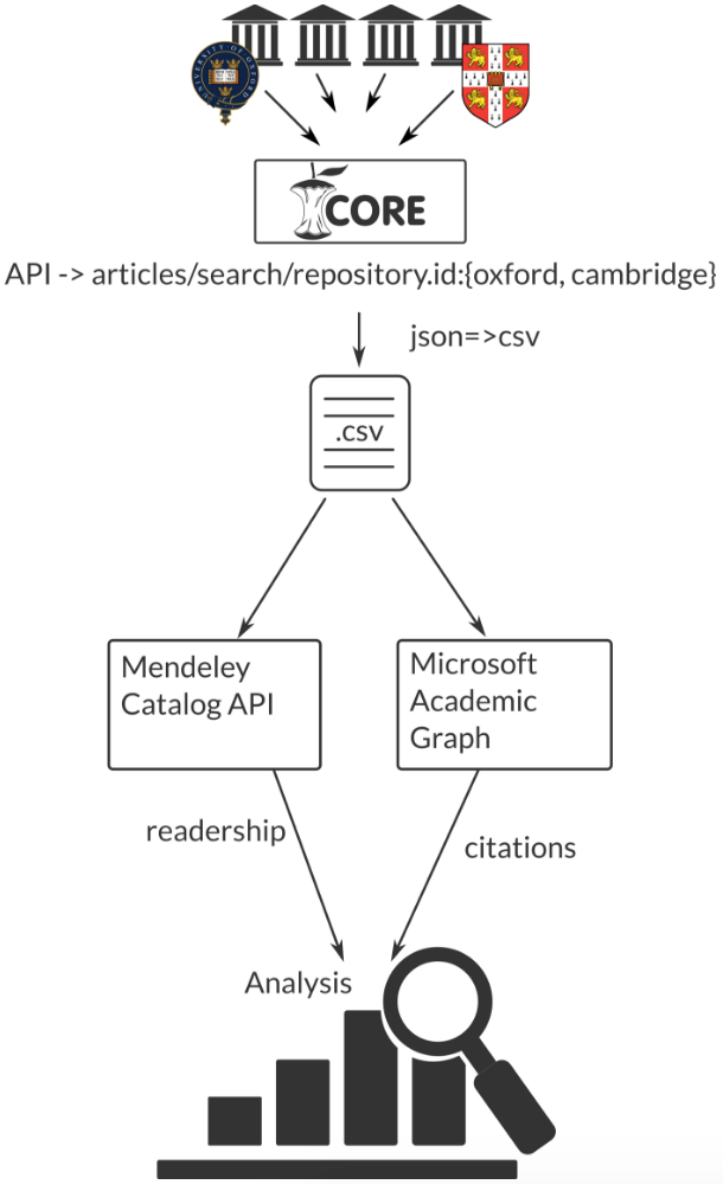
Cambridge vs Oxford Research Impact Contest

Image source: The JeanRichard Aquascope Boat Race
http://www.horbiter.com/en/jeanrichard-aquascope-boat-race/

Workflow

1. CORE harvests the repositories of both institutions. Publication records for a given institution accessible through the CORE API.
2. Microsoft Academic Graph is the world's largest free open citation dataset we use to enrich our data.
3. Mendeley is one of the most popular research reference managers and network. We use it as a free source of "altmetric" information – Mendeley readership.

- We analyse!



CORE

all : Global operations

Show/Hide | List Operations | Expand Operations | Raw

POST	/search	Batch operation for search through all resources
GET	/search/{query}	Search through all resources

articles : Operations for retrieving articles

Show/Hide | List Operations | Expand Operations | Raw

POST	/articles/get	Batch operation for retrieving articles by CORE ID
GET	/articles/get/{coreId}	Get article by CORE ID
GET	/articles/get/{coreId}/download/pdf	Get fulltext PDF by CORE ID
GET	/articles/get/{coreId}/history	Get article history by CORE ID
POST	/articles/search	Batch operation for search through articles
GET	/articles/search/{query}	Search through all documents
POST	/articles/similar	Get articles by similarity to a text

journals : Operations for retrieving journals

Show/Hide | List Operations | Expand Operations | Raw

POST	/journals/get	Batch operation for retrieving journals by ISSN
GET	/journals/get/{issn}	Find journal by ISSN
POST	/journals/search	Batch operation for search through journals
GET	/journals/search/{query}	Search through journals

repositories : Operations for retrieving repositories

Show/Hide | List Operations | Expand Operations | Raw

POST	/repositories/get	Batch operation for retrieving repositories by CORE repository ID
GET	/repositories/get/{repositoryId}	Get repository by CORE repository ID
POST	/repositories/search	Batch operation for searching through repositories
GET	/repositories/search/{query}	Search through all repositories

Step 1. Get publications for a given institution

Call the CORE API to retrieve all articles for a given repository

In [4]:
```
page = 1
oxford_repository_id = 88
r_oxford = CoreApiRequestor(oxford_repository_id, '#api_key_here')
cambridge_repository_id = 27
r_cambridge = CoreApiRequestor(cambridge_repository_id, '#api_key_here')
url = r_oxford.get_request_url(page)
url
```

Out[4]:
```
'http://core.ac.uk/api-v2/articles/search/repositories.id:88?apiKey=#api_key_here&page=1&pageSize=10'
```

We export only the fields we are interested in (title, DOI, etc.) to csv for all records in both repositories

This are the records we get from the CSV for each repository

In [6]:
```
oxford = pandas.read_csv('exports/oxford_export0616.csv',sep=',')
cambridge = pandas.read_csv('exports/cambridge_export0616.csv',sep=',')
print ("Oxford total records: %d Cambridge total records: %d" % (oxford.Title.count(), cambridge.Title.count()))
```

```
Oxford total records: 248865 Cambridge total records: 193249
```

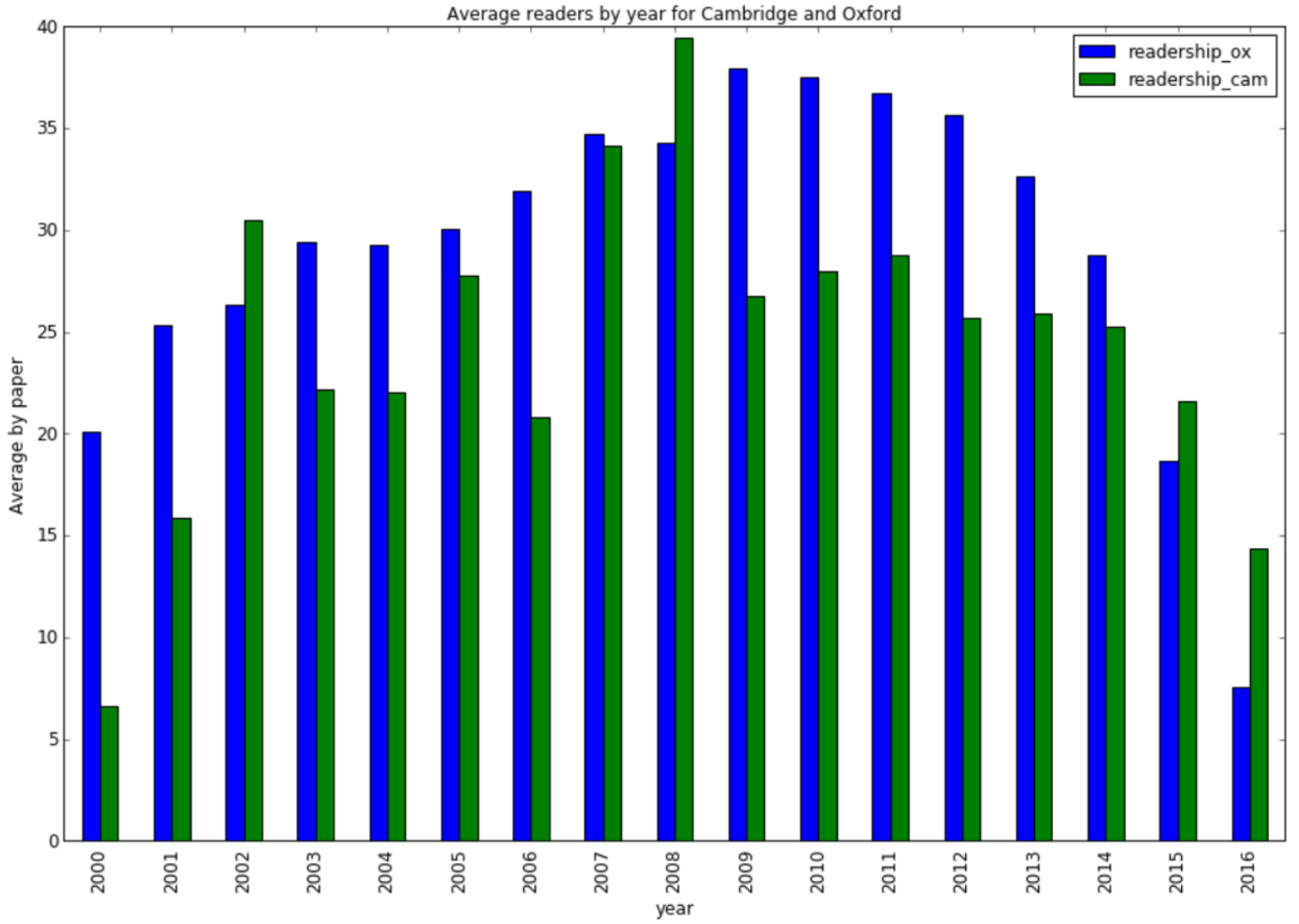
Step 2. Enrich the dataset with Mendeley readership

Calling Mendeley API to get readearship values based on the DOIs

```
def get_readership(mendeley, dataset):
    #...
    doc = mendeley.catalog.by_identifier(doi=value['DOI'], view="stats")
    total_readers.append(doc.reader_count)
    total_readers_by_country.append(doc.reader_count_by_country)
    years.append(doc.year)
    dois.append(np.nan)
    #...
```

	Title	DOI	year	total_readers
196360	Five Misunderstandings About Case-Study Research	10.1177/1077800405284363	2006.0	5344.0
32534	Preferred reporting items for systematic revie...	10.1371/journal.pmed.1000097	2009.0	3245.0
39592	Food security: the challenge of feeding 9 bill...	10.1126/science.1185383	2010.0	2942.0
9158	Power failure: Why small sample size undermine...	10.1038/nrn3475	2013.0	2637.0
3572	Finding the missing heritability of complex di...	10.1038/nature08494	2009.0	2526.0

Average readers per year from 2000 to 2016



Step 3. Enrich with citations from MAG

60279	Preferred reporting items for systematic revie...	10.1371/journal.pmed.1000097	2009	5168
24059	A new statistical method for haplotype reconst...	10.1086/319501	2001	4489
48772	New diagnostic criteria for multiple sclerosis...	10.1002/ana.410130302	1983	4021
22033	Place navigation impaired in rats with hippoca...	10.1038/297681a0	1982	3612
47133	A NOVEL GENE CONTAINING A TRINUCLEOTIDE REPEAT...	10.1016/0092-8674(93)90585-E	1993	3559
27564	Advances in functional and structural MR image...	10.1016/j.neuroimage.2004.07.051	2004	3444
72213	Fast robust automated brain extraction.	10.1002/hbm.10062	2002	3298
59156	Association of glycaemia with macrovascular an...	10.1136/bmj.321.7258.405	2000	3052
43523	Blood pressure, stroke, and coronary heart dis...	10.1016/0140-6736(90)90878-9	1990	2933
475	Finding the missing heritability of complex di...	10.1038/nature08494	2009	2867
20970	Effect of intensive blood-glucose control with...	10.1016/S0140-6736(98)07037-8	1998	2704
13418	The PRISMA statement for reporting systematic ...	10.1136/bmj.b2700	2009	2562
59219	Pressure-temperature-time path discontinuity i...	10.1130/0091-7613(2002)030	2002	2396
23520	Phenotypic Analysis of Antigen-Specific T Lymp...	10.1126/science.274.5284.94	1996	2372
24793	Segregation of a missense mutation in the amyl...	10.1038/349704a0	1991	2211

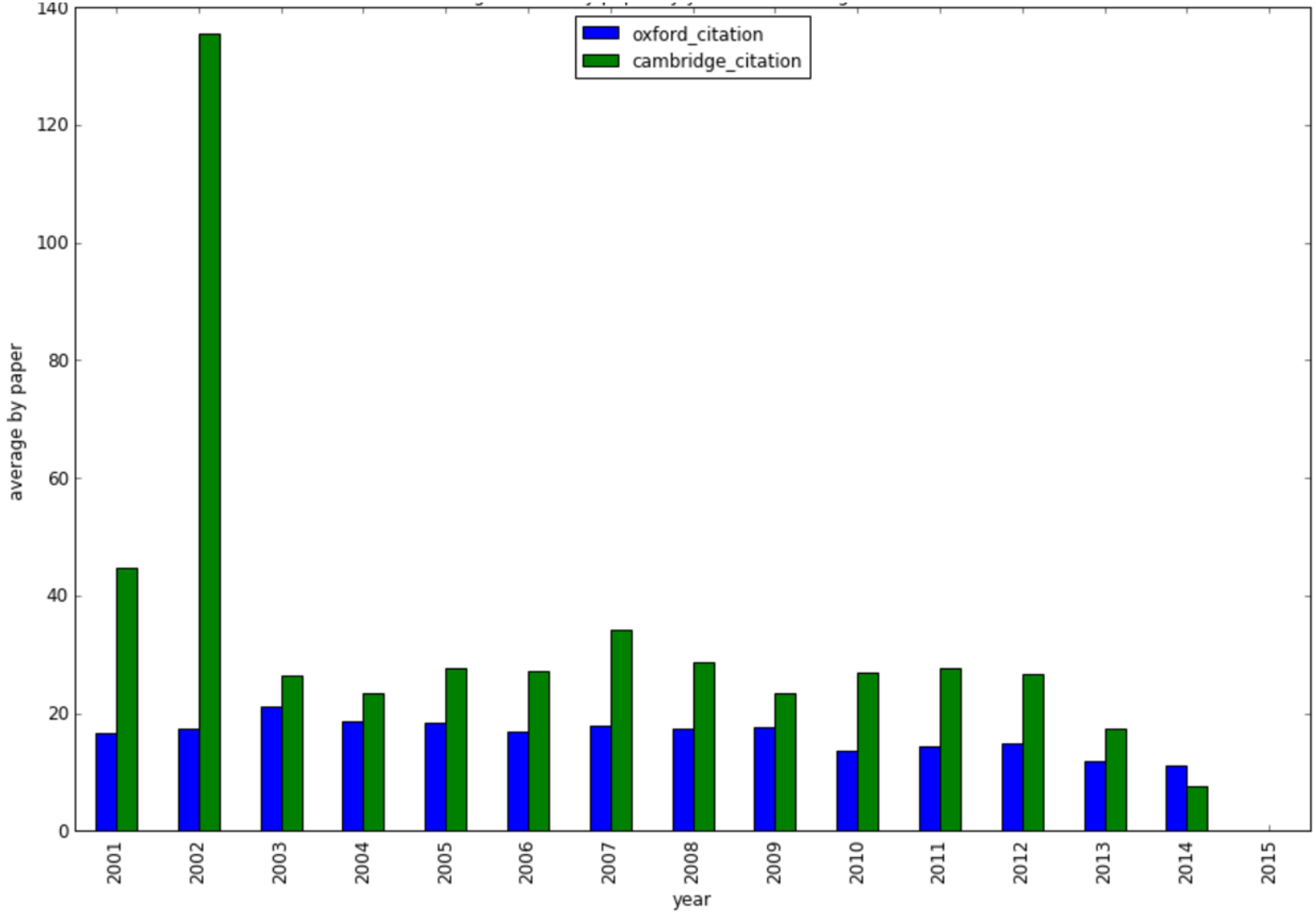
Post process citation data

- Remove papers with 0 citations
- Merge by DOI
- Aggregate by year

Out[10]:

	publish_year	cambridge_citation	oxford_citation
0	2001	121.400000	44.835334
1	2002	357.833333	46.096226
2	2003	67.500000	54.195415
3	2004	58.115385	46.422710
4	2005	66.100000	43.820861
5	2006	62.387097	38.794846
6	2007	74.948276	39.516109
7	2008	59.621212	35.986594
8	2009	45.790323	34.427008
9	2010	48.232143	24.607564
10	2011	44.641379	23.072577
11	2012	36.809524	20.610609
12	2013	19.131148	12.945605
13	2014	5.282474	7.749221
14	2015	2.297778	3.078358

Average citation by paper by year



And the winner is…

Oxford University papers have a **higher readership**

Oxford vs Cambridge

1 – 0

University of Cambridge papers are **cited more often**

Oxford vs Cambridge

0 – 1

CORE Metrics - I

Search About Dashboard Data Providers Applications Blog Internships Contact us

Help **Feedback** My account Logout

Find in CORE

Apollo Edit

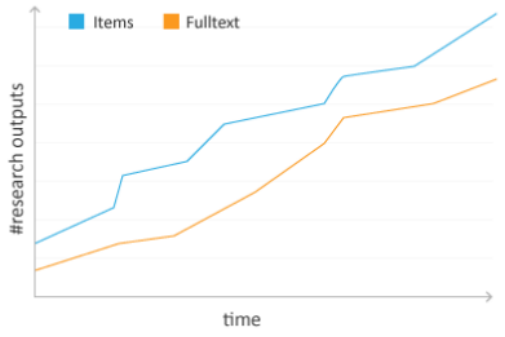
University of Cambridge

support@repository.cam.ac.uk

Profile Content Issues Export RIOXX Compliance IRUS

Content

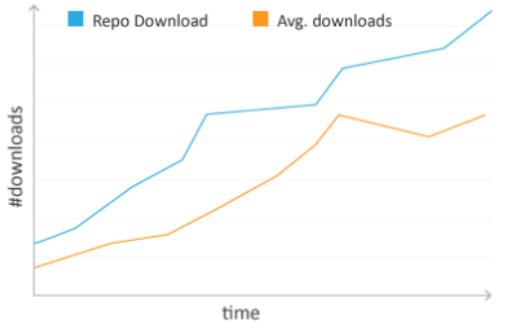
Rank:	5th =
Items:	175.000 ▼
Fulltext:	70.000 ▲



Metrics

Citations:	2.5M =
Citation Rank:	16th ▼
Download:	6Mlm ▲
Downloads rank:	3rd ▲
Readers:	200k ▲

Downloads Readers Citations Semantometrics



CORE Metrics - II

Thank you!

Petr Knoth, Research Fellow, petr.knoth@open.ac.uk

Nancy Pontika, Open Access Aggregation Officer, nancy.pontika@open.ac.uk

Website: http://core.ac.uk
Email: theteam@core.ac.uk
Twitter: @oacore